

On Bayesian Modelling of the Uncertainties in Palaeoclimate Reconstruction

Andrew C. Parnell^{1,2}, James Sweeney³, Thinh K. Doan⁴, Michael Salter-Townshend²,
Judy R.M. Allen⁵, Brian Huntley⁵, and John Haslett⁴

¹School of Mathematical Sciences (Statistics), University College Dublin, Ireland

²Complex and Adaptive Systems Laboratory, University College Dublin, Ireland

³School of Mathematical Sciences, University College Cork, Ireland

⁴School of Computer Science and Statistics, Trinity College Dublin, Ireland

⁵School of Biological and Biomedical Sciences, Durham University, UK

March 8, 2013

Abstract

We outline a model and algorithm to perform inference on the palaeoclimate and palaeoclimate volatility from pollen proxy data. We use a novel multivariate non-linear non-Gaussian state space model consisting of an observation equation linking climate to proxy data and an evolution equation driving climate change over time. The link from climate to proxy data is defined by a pre-calibrated forward model, as developed in Salter-Townshend and Haslett (2012) and Sweeney (2012). Climatic change is represented by a temporally-uncertain Normal-Inverse Gaussian Levy process, being able to capture large jumps in multivariate climate whilst remaining temporally consistent. The pre-calibrated nature of the forward model allows us to cut feedback between the observation and evolution equations and thus integrate out the state variable entirely whilst making minimal simplifying assumptions. A key part of this approach is the creation of mixtures of marginal data posteriors representing the information obtained about climate from each individual time point. Our approach allows for an extremely efficient MCMC algorithm, which we demonstrate with a pollen core from Sluggan Bog, County Antrim, Northern Ireland.

Keywords: Palaeoclimate Reconstructions, Normal-Inverse Gaussian Process, Marginal Data Posteriors, State Space Models, Temporal Uncertainty

1 Introduction

In this paper we aim to answer the following two questions:

- (1) What was climate during period (t_i, t_j) in location s ?
- (2) What was the *volatility* in climate during period (t_i, t_j) in location s ?

For illustration, we set $(t_i, t_j) = (0, 14)$ thousand years before present (written ‘ka BP’¹), and s to be the central part of Northern Ireland. Were t_j to be restricted to a smaller value we could use recent instrumental measurements for which we could obtain a fairly precise answer. When t_j is allowed to grow large, we need to use proxy data: indirect records of climate stored in biological and mineral deposits. In this paper we outline a method to perform inference on the palaeoclimate (sometimes loosely referred to as ‘reconstruction’) and thus answer our questions of interest using pollen data, following on from Haslett et al. (2006). However,

¹We use the common convention in setting the ‘present’ to be 1950AD

our methods are general and readily apply to other proxy data series or even multiple proxies simultaneously.

Palaeoclimate reconstruction is a major focus of the Intergovernmental Panel on Climate Change (Jansen et al., 2007). Public interest, however, has largely been fuelled by the ‘Hockey Stick’ and ‘climategate’ controversies, e.g. Mann et al. (1998, 1999); McShane and Wyner (2011) where (t_i, t_j) was $(0, 1)$ ka BP and s was the Northern Hemisphere. Climate changes during the past millennium are relatively small and can be inferred with reasonable precision from proxies (e.g. tree rings) that are resolved annually. In contrast, the much older Younger Dryas period (12.8ka to 11.5ka BP) shows a rapid switching from warm to cold to warm. During this period ice core data from Greenland show abrupt warmings of up to 16 °C within decades (Jansen et al., 2007, P435). This size and rate of change is not captured well by the General Circulation Models (GCMs) which are used to predict future climate, nor by the precise proxies used to examine the past millennium. Pollen proxy data offer the best hope of resolving such sizeable past climate changes in locations other than Greenland.

The model we propose differs from many recent studies performing inference on palaeoclimate because (a) it is non-linear and non-Gaussian in the relationship between climate and proxy, (b) we infer only climatic variables to which the proxy is sensitive, (c) we use real data to produce climate reconstructions rather than simulated ‘psuedo-proxies’, (d) we take account of chronological uncertainty, and (e) we allow for stochastic variability in climate. This approach, tackling many more sources of uncertainty than previously, is inspired by the SUPRANET group², of which many ideas are shared with Tingley et al. (2012) and Huntley (2012). We expand more on the importance of these various aspects in subsequent sections.

The pollen data we use arise from cores taken beneath a lake or bog. The chosen site will have accumulated many thousands of years of pollen from plants and trees that have surrounded it. After extraction, the core will have been sliced into narrow depths d_i , each treated as a near instantaneous snapshot of the vegetation at that depth. From each slice a palynologist will count many different varieties of pollen and record the counts and depths. We make use of 28 of these counts, corresponding to pollen varieties that have been shown to be representative of a particular aspect or boundary condition on one or more aspects of climate (Huntley, 1993). We write this 28-vector as y_i . The information about climate from the pollen is obscured by: the delay in the reaction of plants and trees to climate change; the ability of the palynologist to recognise the desecrated remains of fossil pollen under a microscope; and the degree to which different species disseminate differentially recognisable pollen. The former of these is considered negligible over the timescales with which we work. The latter two are included as part of our likelihood, being both zero-inflated and highly non-linear in the relationship between climate and pollen.

The practicalities of sampling lake sediment mean that the depths d_i may not be taken regularly throughout the core. Besides, even if the depths in the core are regularly taken, the times to which they correspond depend on the rate of sedimentation and may be very non-linear. The usual approach to transform depth into time is to take between 5 and 20 radiocarbon dates from various depths (again not necessarily regularly), and use these to build a monotonic age-depth model which supplies times t_i from every layer in the core; a number of statistical age-depth models have been proposed (e.g. Bronk Ramsey, 2008; Haslett and Parnell, 2008; Blaauw and Christen, 2011). A side-effect of using one of these models is that times t_i are now uncertain, governed by a draw from a probability distribution $\pi(t|d)$ with t and d representing all of the times and depths in the core. Each draw is termed a chronology.

Taken altogether, we can write our model in state space form as:

$$y_i|c_i \sim f_\theta(c_i, \mathcal{D}), \quad i = 1 \dots, n \quad (1)$$

$$c_i = c_{i-1} + \gamma_i, \quad i = 2, \dots, n \quad (2)$$

²Studying uncertainty in palaeoclimate reconstructions: <http://caitlin-buck.staff.shef.ac.uk/SUPRANet/>

where $y_i = y(t_i)$ represents pollen data at time t_i , $c_i = c(t_i)$ is palaeoclimate, f is a *forward model* parameterised by θ , and $\gamma_i \sim N(0, v(t_{i-1}, t_i))$ are residuals. Following convention, we call Equation 1 the *observation equation*, and Equation 2 the *evolution equation*. All of the parameters involved are multivariate: y_i being here of dimension 28 and c_i being of dimension 3. In our example in Section 6 $n = 115$. The parameters θ and the relationship f are well known and thus pre-calibrated from external data sources (a modern data set of pollen and climate $\mathcal{D} = (c^m, y^m)$), and so our focus is on the posterior distribution $\pi(c, v|y)$. Our particular interest is in the distribution of the incremental variance terms $v_i = v(t_{i-1}, t_i)$, representing the square of the volatility. We set these as being Inverse Gaussian (Barndorff-Nielsen, 1997), written $IG(\eta, \phi)$, yielding an infinitely divisible long-tailed stochastic process on c .

From the age-depth model and the state space equations above we can form a posterior distribution:

$$\pi(c, v, t, \eta, \phi|y, d) \propto \pi(\eta, \phi) \pi(t|d) \pi(c_1) \prod_{i=1}^{n-1} \pi(v_i|\eta, \phi, t) \prod_{i=2}^n \pi(c_i|c_{i-1}, v_i) \prod_{i=1}^n \pi(y_i|c_i) \quad (3)$$

where $\pi(\eta, \phi)$ is the prior distribution on (η, ϕ) , $\pi(t|d)$ is the age-depth model, $\pi(c_1)$ is the marginal prior on the first climate time point (which we treat as flat), $\pi(v_i|\eta, \phi, t)$ is the Inverse Gaussian prior on increment variances, $\pi(c_i|c_{i-1}, v_i)$ is the evolution equation distribution, and $\pi(y_i|c_i)$ is the likelihood from the observation equation distribution.

Our focus in this paper is not on the likelihood $\pi(y_i|c_i)$, the version we use having been discussed in detail by Salter-Townshend and Haslett (2012), and Sweeney (2012). Instead, we focus on how to fit the model defined by Equation 3 in a situation where repeated calls to the likelihood are costly. In fact we separate out the calculation of the likelihood into:

- (a) An initial calibration step, where we cut feedback from θ to form $\pi(y_i|c_i) = \int \pi(y_i|c_i, \theta) \pi(\theta|\mathcal{D}) d\theta$. From these we create *marginal data posteriors* (MDPs) of the form $\pi(c_i|y_i)$.
- (b) A second model-fitting step where we use the MDP in place of the likelihood to create the posterior distribution $\pi(c, v, t, \eta, \phi|y, d)$ as above.

We justify the validity of step (a) in Section 4; the remainder of the paper is concerned with step (b). A Directed Acyclic Graph (DAG) for our complete model is shown in Figure 1 and discussed further in Section 5.

A key aspect of our approach is that many of the parts of Equation 3 are treated as mixtures of Gaussian distributions. The marginal data posteriors are approximated as finite mixtures of Gaussians, which allows the posterior to be factorised, and removes the need for c to be estimated during the model fitting stage. We can then create sample paths of $c(t)$ as a mixture of Gaussian distributions depending on the NIG parameters and the MDPs. Similarly, the NIG prior can be seen as a mean-scale infinite mixture of a Gaussian distribution where the mixing distribution is Inverse Gaussian. We make use of this property to give fast updates of the increment variance parameters v .

Our approach fundamentally differs from previous work in the use and formatting of data used as ‘input’ to an inference procedure. The work of Mann et al. (1998) and Mann et al. (1999) use ‘products’: pre-treated proxy data sets that have been aggregated and transformed on to a scale that is supposed to match an aspect of climate such as mean annual temperature. Often the details of such transformations are lost to the final inference so that the link between proxy data and climate, especially with respect to uncertainty, is not considered during the reconstruction of past climate. Similarly, other authors (e.g. Li et al., 2010) have simulated their proxy data from climate model output, so that the input format is analogous to that of products. By contrast, we use an inference procedure that takes raw pollen data directly to perform inference on climate. However, the marginal data posteriors we produce might be considered as products though with a more fully quantified estimate of uncertainty.

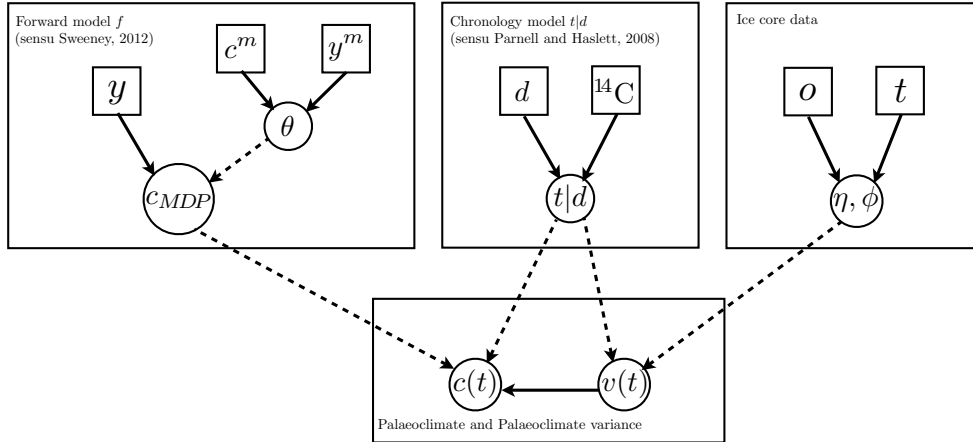


Figure 1: A Directed Acyclic Graph for our palaeoclimate model. Broken lines indicate relationships where feedback is cut. See Section 5 for more details on the modelling relationships and Technical Appendix A for details on the notation.

The paper is organised as follows. In Section 2 we examine previously used methods for creating palaeoclimate reconstructions. In Section 3 we present a simple version of our time series model based on ice core data. In Section 4 we introduce the mathematical justification of our method, and apply it to the palaeoclimate scenario in Section 5. In Section 6 we show the results of our model applied to a site in Northern Ireland. We examine some shortcomings of our model and discuss future directions in Section 7. The mathematical derivation is somewhat complex so we include a notational glossary in Technical Appendix A. We include some of the more technical details of the algorithm in Technical Appendix B. We perform model validation checks in Appendix C. The resulting model is implemented in the R package *Bclim*, which is available online at the R repository (R Foundation For Statistical Computing Austria, 2011). Instructions for using the R package can be found in Technical Appendix D.

2 Previous work in statistical palaeoclimatology

Here we explore the ways in which other authors have dealt with various aspects of palaeoclimate reconstruction. Whilst recent work has focussed on model-based approaches (particularly Bayesian), authors in the more distant past have used other methods. The most famous of these is the Mann et al. (1998, 1999) papers using principal components regression or the regularised EM algorithm, though others such as Huntley (1993) and Ter Braak (1995) follow differing non-model based criteria. Other projects have focussed on certain aspects of palaeoclimate in relation to General Circulation Models (GCMs) or at particular points in time, such as PMIP (Braconnot and Harrison, 2003), and the PALAEOQUMP project (Masson-Delmotte et al., 2006). For brevity, we discuss in detail here only the statistical model-based approaches used by Haslett et al. (2006); Tingley and Huybers (2010); Li et al. (2010) (hereafter H06, T10 and L10 respectively) and focus on the key climatological issues: choice of climate variables, likelihood choices, prior distributions on climate, chronological modelling; and the most relevant statistical issues: fitting models with complex likelihoods, the NIG process, and the treatment of mixture models. A more thorough review of different approaches can be found in Tingley et al. (2012).

2.1 Issues of climatology

A key modelling choice in palaeoclimate reconstruction is the aspects of climate to reconstruct; climate being defined as the ‘average’ of weather. Many authors (e.g. Mann et al., 1998, L10) have reconstructed mean annual temperatures, often over large spatial scales such as the Northern Hemisphere. This may seem like a logical choice, given that these are relatively easy to measure and relevant to human wellbeing. However, as recently summarised by Huntley (2012), the choice should instead depend on the sensitivity of the proxy data to changes in climate. Many biological proxies have shown that they are not sensitive to changes in mean annual temperature so reconstructions of this sort will be biased and overly uncertain. Instead, more relevant variables were proposed by, amongst others, Huntley (1993):

- the Mean Temperature of the Coldest Month (MTCO), a measure of the harshness of the winter,
- the Growing Degree Days above 5°C (GDD5, also known as the annual temperature sum above 5°C), a measure of the length of the growing season,
- the ratio of Actual to Potential Evapotranspiration (AET/PET), a measure of the proportion of available moisture.

The first two of these were used for reconstruction by H06. Even then, however, missing a single important climate variable (in this case concerning moisture) will increase uncertainty, bias and possibly invalidate the conditional independence of the likelihood. Other non-biological proxies (e.g. Borehole temperatures, Brynjarsdóttir and Berliner, 2011) have shown sensitivity to surface temperature of only the surrounding microclimate. The choice of climate variables is even more important in multi-proxy reconstructions, where the differing sensitivities and temporal scales at which proxies are formed needs to be carefully understood before models can be created.

Once the climate variables have been chosen, we can posit a likelihood $\pi(y|c, \theta)$. We see the likelihood here as a data generating mechanism providing realistic proxy data given the relevant aspects of climate, often also called a forward model indicating the direction of causation (from climate *to* proxy). Numerous methods for creating a link between climate and pollen proxy data have been suggested, see Ohlwein and Wahl (2012) for a full review. There are two approaches which stand out as being most popular. The first is to perform a calibration via modern analogue data, as used by H06, T10 and L10. The idea is to find data, here \mathcal{D} , for which both precise proxy and climate are available, and upon which the likelihood can be trained. The version used by T10, L10 requires the calibration data to be temporally overlapping, through which a temporal proxy/climate relationship might be inferred. The version used by H06 has no temporal overlap, but rather uses the climatological location of modern data in a multivariate psuedo-spatial model where the proxy is the response. This model has been substantially enhanced in Salter-Townshend and Haslett (2012) and Sweeney (2012), and the latter is the version we use for our palaeoclimate model. The second approach is to use a physical model, as employed by Garreta et al. (2009) with mixed success. The physical model can provide a far richer relationship, taking account of differences in pollination mechanism and productivity between trees, and their response to complex inter-relations between climate variables. However, the models used are often extremely slow and may lack a stochastic element, making their use in Equation 3 difficult.

In either scenario a likelihood is created such that:

$$y_i|c_i, \theta \sim f(h(c_i, \theta_h), \theta_\epsilon) \quad (4)$$

where h is a function which accounts for the sensitivity of the proxy to climate with parameters θ_h , and measurement error captured by θ_ϵ . Together $\theta = (\theta_h, \theta_\epsilon)$. Both parameters and their relationship through h are estimated from the calibration data, \mathcal{D} . In Sweeney (2012) this relationship is explicit; f is a zero-inflated Negative Binomial distribution, with θ_h the parameters of a Gaussian Markov Random Field within a nested Dirichlet distribution specifying the proportions of each taxa. In L10 and T10 f is Gaussian (the proxy data y are ‘products’) with h a linear regression so that θ_h represents regression parameters

and θ_ϵ residual variances and covariances. Both L10 and T10 use simulated data (‘pseudo proxies’) so it is not stated exactly which transformation of y might link back to a real-world biological proxy such as pollen.

In H06 the prior distribution on climate was a t_8 random walk which is not closed under addition. Both L10 and T10 use more conventional time series methods, employing a Gaussian MA(2) and AR(2) model respectively. In T10 the prior distribution is allowed to vary spatially with a stationary Exponential covariance structure. In L10 the prior is expanded to include extra covariates corresponding to known forcings in a linear fashion. Unfortunately these structures are not applicable to the present situation where time is uncertain and continuous. In palaeoclimate reconstruction from pollen, uncertainty in time arises from radiocarbon dates taken from the pollen core which require non-linear calibration (see Buck et al., 1996). These contrast with the other proxy data used by L10 and T10 such as tree rings for which the times are known precisely. For our purposes we use the Bchron software of Haslett and Parnell (2008), being explicitly built for inclusion in palaeoclimate reconstructions.

2.2 Issues of statistical modelling

When the likelihood and prior distributions are Gaussian, as in T10 and L10, then Equation 3 is conceptually easy to solve via traditional MCMC or Particle-type methods. Things are complicated slightly by the high-dimensionality of the climate and volatility parameters and the sheer quantity of data. H06, however, have a non-linear model and thus combine the MCMC methods with some approximations which cut feedback between parts of the model. In particular, parameters θ of the likelihood are learnt solely from the modern data. Thus information from any particular core contributes no ‘further’ information. In this paper we use a likelihood which is considerably more complex than that of H06, and is thus far too slow to solve using these methods.

A number of alternatives have been proposed for fitting models with complex likelihoods, including but not limited to: Approximate Bayesian Computation (ABC; Beaumont et al., 2002), which relies solely on simulations from the likelihood and user-specified sufficient statistics to compute likelihood ‘distances’; and statistical emulation (SE; Kennedy and O’Hagan, 2001), which builds a stochastic approximation to the likelihood through a designed experiment. However, since we have reasonably good access to a well-defined likelihood it seems inappropriate to throw away much of the information in an ABC type algorithm. Similarly, whilst SE has been shown to work well in a high dimension climatological situation (Holden et al., 2010), it is mostly used for deterministic models (though see Henderson et al., 2010).

As stated earlier, our model can be usefully seen in multivariate state-space terms for which there are numerous algorithms based on Sequential Monte Carlo (see, e.g. Carvalho et al., 2010, for a review). These proceed (in our notation) in a ‘forward’ or filtering stage³ by simulating from $\pi(c_1)$ and then forming $\pi(c_k|y_1, \dots, y_k)$ sequentially for $k = 2, \dots, n$. The filtering densities $\pi(c_k|y_1, \dots, y_k)$ are related to our marginal data posteriors, being posterior distributions on a subset of the data. However, their creation requires both the use of the evolution and the state equation for every time point. In the subsequent sections, we show that in our situation it is possible to produce a valid joint posterior via only the MDPs, which does not require calculation of the evolution density at the forward stage.

A related problem in the estimation of state space models is known as the smoothing or ‘backwards’ step to produce a full joint posterior $\pi(c_1, \dots, c_n|y_1, \dots, y_n)$. It is well-known that this is a challenging problem for particle methods (Doucet et al., 2008), especially in the case of large n and where inference on other parameters are also required. Indeed, our focus is on this smoothing stage of state-space modelling, though even this step (when using the nomenclature of particle filters) over-stresses the focus on the posterior mean and is less natural when the focus is, as here, on past volatility. By contrast, we are in a somewhat unusual situation in that we are happy to cut feedback so that the likelihood depends only on the state parameters

³Unrelated to the forward model f presented in Equation 1.

(climate c) with the distribution on θ known. When the prior is intrinsic we show that it is possible to integrate out our state variable c and bypass many of the issues caused by particle degeneracy and sample size. We thus believe that our MDP approach is superior (achieving faster convergence even for large n), though we report elsewhere on the general properties of our fitting algorithm.

We use the NIG as a prior distribution on differences $c_i - c_{i-1}$ across continuous time. The distribution was introduced by Barndorff-Nielsen (1997) as a mean-scale mixture of Gaussian distributions. If $x_i|v_i \sim N(\mu + \beta v_i, v_i)$ and $v_i \sim IG(\alpha, \delta)$ where IG is in the Inverse Gaussian distribution (Betrò and Rotondi, 1991), then $x_i \sim NIG(\mu, \beta, \alpha, \delta)$ with mean μ , skew β , and variance/tail parameters α, δ . Most importantly, the NIG distribution is closed under addition, so that if $X_i \sim NIG(\mu_i, \beta, \alpha, \delta_i)$ then $\sum_i X_i \sim NIG(\sum \mu_i, \beta, \alpha, \sum \delta_i)$. Bayesian inference for the NIG distribution has been discussed by Karlis and Lillstol (2004), providing closed-form complete conditionals and thus a neat Gibbs algorithm. Some of these complete conditionals transfer over to our model and are discussed further in Section 3. The form we use is that of the NIG process as discussed by Ribeiro (2003), representing the random walk as a subordinated Brownian motion. The resulting NIG process is, of course, not the only long-tailed smoothing model available, forming part of the larger class of Levy processes (e.g. Barndorff-Nielsen and Shephard, 2001). Similarly, Fonseca and Steel (2011) provide a class of long-tailed smoothers for space-time models. Where discontinuities are expected, a change-point model might be used (e.g Wyse et al., 2010). However, the NIG process is extremely simple to work with, and provides many of the features we might expect to appear in a dynamical system such as climate.

The finite mixture models of which we make use are taken from the Mclust algorithm inspired by Fraley and Raftery (2002). Here, a multivariate random variable x arises as a mixture of G normal distributions with unknown means and variances, so that $x = \sum_{g=1}^G p_g N(\mu_g, \tau_g)$ where the p_g are mixture proportions. We use this technique to approximate our MDPs as mixtures of Gaussians via the EM algorithm (Dempster et al., 1977) which provides posterior modes for μ_g, τ_g . A full summary of the methods by which this type of model may be fitted can be found in Frühwirth-Schnatter (2006). An infinite mixture version via the Dirichlet Process (e.g. Congdon, 2001) would also be appropriate were we to be interested in simultaneously estimating the number of mixture components G .

3 A motivating example: the GISP 2 ice core

We illustrate our long-tailed smoothing approach and how we might answer our motivating questions in a simplified scenario based on the GISP2 ice core of Stuiver (2000). A similar but more sophisticated analysis has been undertaken by Davidsen (2010). The proxy data here are measurements of $\delta^{18}\text{O}$, the deviation from a standard of the ratio of the two stable isotopes of oxygen measured in laminated (layered), and thus precisely dated, ice. However, the data are not observed at regular time intervals. For illustration we assume an identity relationship between $\delta^{18}\text{O}$ and summer temperatures may be appropriate. A plot of the data covering the previous 120ka BP is shown in Figure 2. We write the response variable here as $o_i = o(t_i)$, and fit a model so that:

$$o_i - o_{i-1} \sim N(\mu\Delta_i + \beta v_i, v_i), v_i \sim IG_2(\eta\Delta_i, \phi\Delta_i), i = 2, \dots, n \quad (5)$$

where $\Delta_i = t_i - t_{i-1}$ and IG_2 is the alternative parameterisation of the Inverse Gaussian distribution given by Betrò and Rotondi (1991). Under this parameterisation, the parameter η can be seen as the mean unit-time variance v of the resulting NIG distribution, with parameter ϕ controlling the variance of v . Following Karlis and Lillstol (2004), we set prior distributions so that $\beta \sim N(0, \tau_\beta^{-1})$, $\mu \sim N(0, \tau_\mu^{-1})$, $\eta \sim Ga(a_\eta, b_\eta)$ and $\phi \sim Ga(a_\phi, b_\phi)$, where Ga is a Gamma distribution. An MCMC algorithm now has Gibbs' steps with

the following complete conditionals:

$$\begin{aligned}
\pi(\eta|\dots) &\sim GIG\left(\frac{n-1}{2} + a_\eta, \phi u_1, \phi u_3 + 2b_\eta\right), \\
\pi(\phi|\dots) &\sim Ga\left(\frac{n-1}{2} + a_\phi, \frac{u_1}{2\eta} - u_2 + \frac{u_3\eta}{2}\right), \\
\pi(v_i|\dots) &\sim GIG\left(-1, (o_i - \mu\Delta_i)^2 + \phi\eta(\Delta_i)^2, (\beta\Delta_i)^2 + \frac{\phi}{\eta}\right), \\
\pi(\mu, \beta|\dots) &\sim N\left(\left[\begin{array}{cc} \sum(\Delta_i)^2 v_i + \tau_\mu & \sum \Delta_i v_i^2 \\ \sum \Delta_i v_i^2 & \sum v_i^2 + \tau_\beta \end{array}\right]^{-1} \left[\begin{array}{c} \sum \Delta_i v_i o_i \\ \sum v_i^2 o_i \end{array}\right], \left[\begin{array}{cc} \sum(\Delta_i)^2 v_i + \tau_\mu & \sum \Delta_i v_i^2 \\ \sum \Delta_i v_i^2 & \sum v_i^2 + \tau_\beta \end{array}\right]^{-1}\right)
\end{aligned}$$

Here GIG is the Generalised Inverse Gaussian distribution, and $u_1 = \sum v_i$, $u_2 = \sum \Delta_i - b_\phi$, $u_3 = \sum (\Delta_i)^2 / v_i$. Such a model is extremely quick and easy to fit and, from the resulting estimates of v , we can use an Inverse Gaussian bridge to obtain decadal estimates of climate volatilities. Writing $v(t_i, t_j)$ as the squared volatility associated with time increment (t_i, t_j) , the bridge result for time t^* with $t_i \leq t^* \leq t_j$ is provided by Ribeiro (2003) and gives:

$$\pi(v(t_i, t^*)|v(t_i, t^*) + v(t^*, t_j) = v) \propto v(t_i, t^*)(v - v(t_i, t^*)) \exp\left[-\frac{\phi\mu}{2} \left\{\frac{(t^* - t_i)^2}{v(t_i, t^*)} + \frac{(t_j - t^*)^2}{v - v(t_i, t^*)}\right\}\right], \quad (6)$$

where the quantity inside the exponent is a χ_1^2 distribution. Exact methods for sampling from Equation 6 are outlined in Ribeiro (2003). From the resulting estimates, we can look at the changing volatility in climate over time. Alternatively a return level plot could be created (see e.g. Coles, 2001). Figure 3 shows the posterior distributions of (μ, β, η, ϕ) and the 95% credible intervals for volatilities \sqrt{v} from fitting the above model to the ice core data after a run of 100,000 iterations, discarding 20,000 as burn-in, and thinning by 80, with all hyper-parameters $(a_\eta, b_\eta, a_\mu, b_\mu, \tau_\mu, \tau_\beta)$ set to 0.01, indicating only vague prior knowledge.

4 Inference for marginal-data and joint-data posteriors

In this section we distinguish between the marginal-data posterior $\pi(c_i|y_i)$ and the joint-data posterior $\pi(c|y)$, and subsequently introduce one key aspect of the modelling, finite Gaussian mixtures, that allows us to retain rich models and yet to achieve considerable algorithmic speed. We extend the ideas of the previous section, so that the NIG process is applied to climates c which are now multivariate latent parameters, connected to the proxy data y via a likelihood $\pi(y|c) = \int \pi(y|\theta, c)\pi(\theta|\mathcal{D}) d\theta$. The key benefit of using mixtures of MDPs is that the highest dimensional parameter (climate c) can be simply marginalised out of the model and re-created at a second stage, focussing inference only on the climate volatility parameters.

It is first important to note the distinction between $\pi(c_i|y)$ and $\pi(c_i|y_i)$. The former is the marginalised *joint* data posterior $\pi(c, v, \eta, \phi|y)$ (the left side of Equation 3); the latter, our MDP, represents only the information in c_i obtained from the proxy data y_i at that same layer i . It is trivial to note that the MDP can be obtained directly from the application of Bayes' theorem to the conditionally independent likelihood:

$$\prod_i \pi(y_i|c_i) \propto \frac{\prod_i \pi(c_i|y_i)}{\prod_i \pi(c_i)} \quad (7)$$

There are two features of our model which make the use of such MDPs feasible. First, we are happy to cut feedback so that the parameters θ are integrated out of the likelihood. Second the prior on climate is intrinsic so, provided that a flat prior is applied to $\pi(c_1)$, all other $\pi(c_j)$ are flat too, thus the denominator of Equation 7 is a constant. Now the creation of $\pi(c_i|y_i)$ for every layer can be achieved in parallel at an offline early step of the process, with the lookup of $\pi(c_i|y_i)$ now being near-instantaneous. The

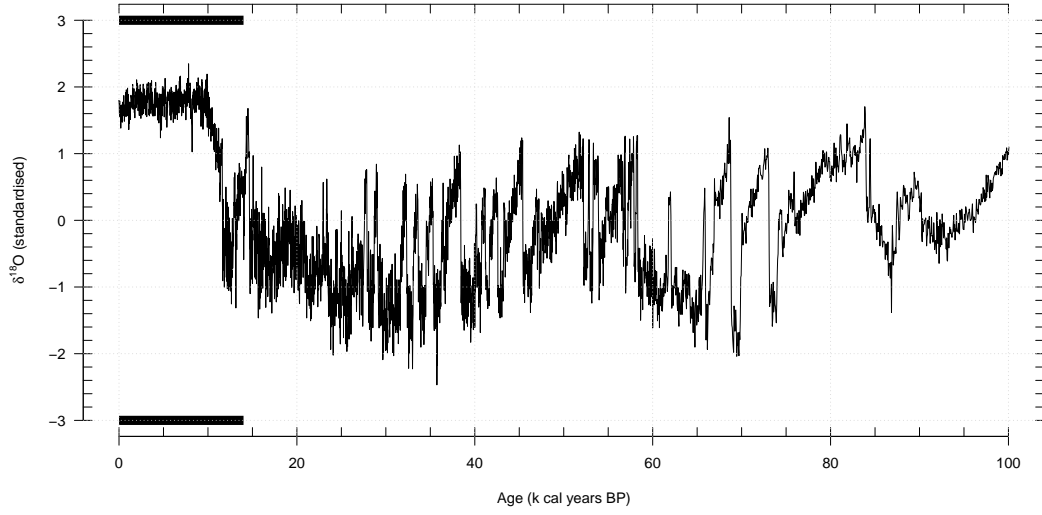


Figure 2: The GISP2 ice core data of Stuiver (2000). The y -axis contains the stable isotope measurement $\delta^{18}\text{O}$ after standardisation to have mean 0 and standard deviation 1. A high value corresponds approximately to warmer summer temperatures, and the lower to colder summer temperatures. The period 0 to 14ka BP is highlighted to match with Figure 3 (right panel) which shows the volatilities from this period.

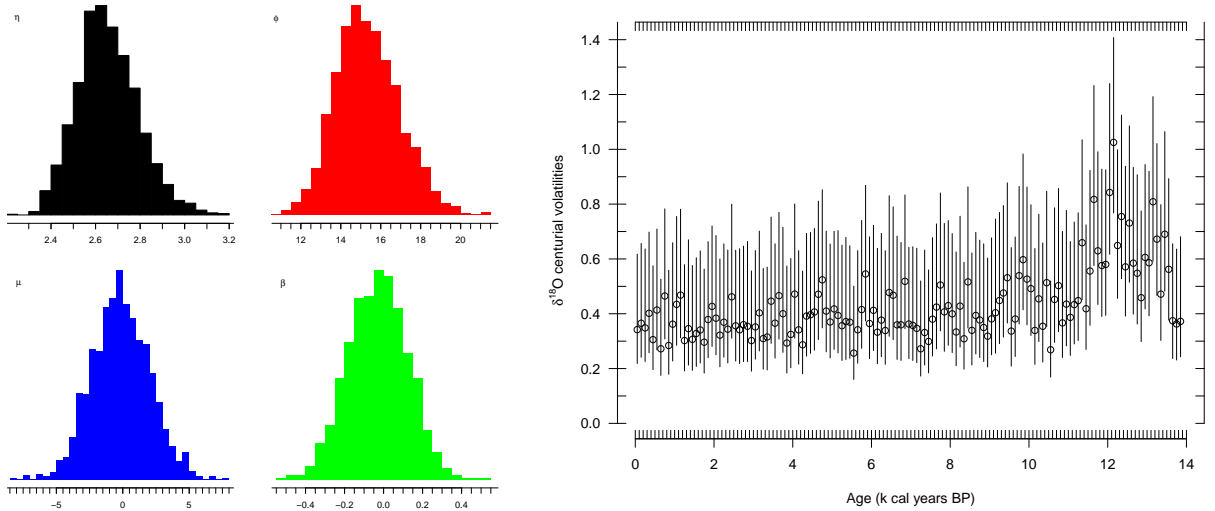


Figure 3: Posterior plots from fitting the model defined by Equation 5 to the GISP2 ice core data. The left plots show the posterior distributions of parameters μ , β , η and ϕ , whilst the right-most plot shows the posterior distribution of v over the time interval 0 to 14ka BP. The empty circles indicate the posterior mean whilst the vertical lines indicate the 95% credibility intervals. The century with the highest (posterior mean) climate volatility is 12.1k to 12.2k BP.

‘cutting feedback’ from likelihood parameters is an unremarkable and implicit assumption in most studies involving instrumental data; e.g., experimental data on temperature are not normally used to re-calibrate the thermometer that generated them. It is interesting to note that in many branches of applied statistics what is considered ‘data’ might actually be an MDP, though Equation 7 would never be explicitly considered.

More formally, we now consider replacing the likelihood in Equation 3 with the *product MDP* $\prod_i \pi(c_i|y_i)$ which we write as $\pi_{\text{MDP}}(c|y)$ to avoid confusion with the marginalised joint data posterior $\pi(c|y)$. Assuming the above requirements are met (i.e. intrinsic prior on c , and feedback cut between likelihood parameters θ and y) there are several further advantages to performing inference with the MDP:

- Once $\pi(c_i|y_i)$ has been created for every i , we never have to make another call to a potentially expensive product likelihood $\prod_{i=1}^n \pi(y_i|c_i)$.
- c is of lower dimension than y , and thus is easier to store in large quantities.
- Different methods for creating the marginal-data posterior can be compared without any need to re-run the full likelihood computation. Although in this paper we use MDPs constructed by our research group, the methodology we describe can in principle receive input from any Bayesian procedure.
- We can explore marginal-data posterior distributions, and remove them at will to cross-validate.

In fact, such an approach may have advantages in other areas of Bayesian statistics. Clearly the approach does not require a flat prior on c_1 , although that is what we use: any prior distribution in which the joint and marginal prior distributions are available will yield a suitable joint posterior. We now elaborate with a simple case in which the likelihood, and thus also the MDP, is Gaussian.

4.1 A simple example

For illustration we consider a simple case with fixed equal time steps and a simple Gaussian random walk prior $c_i - c_{i-1} \sim N(0, v)$. We assume a likelihood $y_i|c_i \sim N(c_i, \tau_i^{-1})$ with known τ_i so the MDP is trivially $c_i|y_i \sim N(y_i, \tau_i^{-1})$. The target of the estimation is $\pi(c, v|y)$. Some simple algebra shows that the posterior factorises: $\pi(c, v|y) \propto \pi_{\text{MDP}}(c|y)\pi(c|v) = \pi(c|y, v)\pi(y|v)\pi(v)$. The distribution $\pi(c|y, v)$ here is Gaussian: $c|y, v \sim N(c; VDy, V)$. The distribution $\pi(y|v)$ is also Gaussian on first differences of y (though this does not extend to the more complicated examples we present in the next section). Here D is the diagonal precision matrix with known components τ_i and $V = (D + W)^{-1}$. V in turn involves the precision matrix W of the random walk, being $v^{-1}B^TB$ where B is an $(n-1) \times n$ differencing matrix with the first row structured as $(-1, 1, 0, \dots, 0)$, and subsequent rows structured similarly. Note that despite the prior on c being improper all matrices involved in $\pi(c|y, v)$ and $\pi(y|v)$ are of full rank. The precision matrix $(D + W)$ is itself tri-diagonal and so allows fast inversion. Taken altogether, we can design a fast algorithm for computing the posterior by first performing inference on the parameter v (by marginalising over c) and subsequently creating a posterior for c via $\pi(c|y, v)$.

4.2 Multivariate Finite Mixtures for MDPs

The model above can be extended to the non-linear case where the MDP is $c_i|y_i \sim N(\mu_i = \mu(y_i), d_i^{-1} = d(y_i)^{-1})$. Following the same approach we obtain the joint data posterior $\pi(c, v|y) \propto \pi(c|y, v)\pi(y|v)$, where again $c|y, v \sim N(c, VD\mu, V)$. Now, however, $\pi(y|v)$ is non-Gaussian, but the same marginalisation over c is possible. We can further remove the assumption that the marginal data posterior is Gaussian by extending to the case where the MDP is treated as a finite mixture. More fully, we now consider MDPs which can be written, for an m -dimensional c_i , as $\pi(c_i|y_i) = \sum_{g=1}^G p_{ig} N(\mathbf{c}_i; \boldsymbol{\mu}_{ig}, \boldsymbol{\tau}_{ig}^{-1})$ where we now use bold type to clearly indicate the use of multivariate climate at each layer i . The terms p_{ig} are probabilities and $(\boldsymbol{\mu}_{ig}, \boldsymbol{\tau}_{ig}^{-1})$ are vector and matrix respectively; these are all functions of the count data \mathbf{y}_i associated with

layer i . In the previous section, there was but one component in each mixture and we identified the sole μ_i with a scalar y_i .

The formulation may alternatively be expressed in terms of an underlying indicator random variable k_i , which with probability p_{ig} selects component g . Conditional on k_i , the MDP is simply Gaussian, and we have but a small extension of the previous section. We propose this as an approximation to the true MDPs. In the Bayesian situations we envision, the $\pi(\mathbf{c}_i|\mathbf{y}_i)$ are typically only available through (Markov Chain) Monte Carlo samples. The mixtures are then derived by fitting a finite Gaussian mixture to each sample. As detailed in Section 2, we use the MClust algorithm of Fraley and Raftery (2002). This is a once-only exercise; there is no inference in the following on any of these parameters, which are taken to be known. The implications of the approximation are that we have access to the very simple results in the previous section.

It follows that

$$\pi_{\text{MDP}}(\mathbf{c}|\mathbf{y}) = \prod_i \pi(\mathbf{c}_i|\mathbf{y}_i) = \prod_i \left\{ \sum_g p_{ig} N(\mathbf{c}_{ig}; \boldsymbol{\mu}_{ig}, \boldsymbol{\tau}_{ig}^{-1}) \right\} = \sum_{\mathbf{K}} p_{\mathbf{K}} N(\mathbf{c}; \boldsymbol{\mu}_{\mathbf{K}}, \boldsymbol{\tau}_{\mathbf{K}}^{-1}) \quad (8)$$

Here the index \mathbf{K} runs over n-tuples $\{k_i; i = 1, \dots, n\}$ and $p_{\mathbf{K}} = \prod_{i,g_i} p_{ig}$. If each mixture involves G components, then there are G^n versions of \mathbf{K} ; the vectors $\boldsymbol{\mu}_{\mathbf{K}}$ hold the appropriate stacked $\boldsymbol{\mu}_{ig}$ and the block diagonal matrices $\boldsymbol{\tau}_{\mathbf{K}}$ have blocks $\boldsymbol{\tau}_{ig}$. We abuse notation somewhat for simplicity, allowing the subscripts (ig) and \mathbf{K} to differentiate the very differently dimensioned vectors ($\boldsymbol{\mu}_{ig}, \boldsymbol{\mu}_{\mathbf{K}}$) and matrices ($\boldsymbol{\tau}_{ig}, \boldsymbol{\tau}_{\mathbf{K}}$). Again \mathbf{K} is an index vector, with associated probability $p_{\mathbf{K}}$. All terms are known. It is to be recalled that the above refers simply to the product of MDPs; it is not the joint distribution $\pi(\mathbf{c}|\mathbf{y})$ that we seek here.

In the following sections we replace the likelihood corresponding to the observation equation with the product MDP represented as mixtures of Gaussians. By integrating out \mathbf{c} we thus bypass the evolution equation and are able to focus our inference on the volatilities at a first stage, updating them via standard Metropolis Hastings MCMC. Subsequently, we can produce climates \mathbf{c} via simple mixtures of multivariate normal distributions defined by the MDPs and the posterior distribution of the volatilities.

5 Palaeoclimate modelling using MDPs

To complete our palaeoclimate model, we combine the MDP approach presented above, which allows us to marginalise over climate \mathbf{c} , with the NIG prior used in Section 3. The NIG prior is applied to each climate dimension individually, so that the time series are treated as independent *a priori*, though dependence is induced through the MDPs which capture relationships between the climate variables. At this stage we also include our age-depth model which allows for uncertainty in the chronology measurement. We use Bchron (Haslett and Parnell, 2008) to provide very many samples of t_1, \dots, t_n , each with the monotone restriction $t_i - t_{i-1} > 0$. In fact, the Compound Poisson Gamma Process used in Bchron could be considered another intrinsic prior and so might be updated as part of our model. However, we do not explore that approach further here, instead preferring to again cut feedback between the age-depth model and the other parameters.

In summary, the following analysis stages are required to perform inference on the joint posterior distribution of climate and volatility:

1. Creation of a suitable chronology via, e.g. Bchron.
2. (P) Creation of MDPs for each layer in the core.
3. (P) Approximation of MDPs as Gaussian mixtures.
4. MCMC run to estimate volatilities \mathbf{v} .

5. Posterior creation of climate \mathbf{c}

6. Interpolation of volatilities \mathbf{v} and climates \mathbf{c} on a pre-chosen grid

This section outlines steps 4, 5 and 6. The inclusion of (P) indicates that this stage can be performed in parallel. The full set of parametric relationships are shown in the DAG of Figure 1.

Our final model can be written as:

$$\mathbf{c}(t_i)|\mathbf{y}_i \sim \sum_{g=1}^G p_{ig} N(\boldsymbol{\mu}_{ig}, \boldsymbol{\tau}_{ig}^{-1}), \quad c_j(t_i) - c_j(t_{i-1}) \sim N(0, v_{ij}), \quad v_{ij} \sim IG_2(\eta_j \Delta_i, \phi_j \Delta_i), \quad (9)$$

where all notation is as before except for the subscript j denoting climate dimension from 1 to $m = 3$. We require the posterior:

$$\begin{aligned} \pi(\mathbf{c}, \mathbf{v}, \boldsymbol{\eta}, \boldsymbol{\phi}, \mathbf{K}, \mathbf{t}|\mathbf{y}, \mathbf{d}) &\propto \left[\prod_{j=1}^m \pi(\eta_j) \pi(\phi_j) \right] \times \pi(\mathbf{K}) \times \pi(\mathbf{t}|\mathbf{d}) \times \left[\prod_{j=1}^m \prod_{i=1}^{n-1} \pi(v_{ij}|\eta_j, \phi_j, \Delta_i) \right] \\ &\times \left[\prod_{j=1}^m \prod_{i=1}^{n-1} \pi(c_j(t_{i+1})|c_j(t_i), v_{ij}) \right] \left[\prod_{i=1}^n \pi_{\text{MDP}}(\mathbf{c}(t_i)|\mathbf{y}_i, k_i) \right] \end{aligned} \quad (10)$$

After some algebra (similar to that in Section 4) we obtain:

$$\pi(\mathbf{c}, \mathbf{v}, \boldsymbol{\eta}, \boldsymbol{\phi}, \mathbf{K}, \mathbf{t}|\mathbf{y}, \mathbf{d}) \propto A \prod_{j=1}^m \frac{N(\mathbf{c}_j; \mathbf{V}_j \mathbf{K} \mathbf{D}_j \mathbf{K} \boldsymbol{\mu}_j \mathbf{K}, \mathbf{V}_j \mathbf{K}) N(\mathbf{0}; \boldsymbol{\mu}_j \mathbf{K}, \mathbf{D}_j^{-1} \mathbf{K})}{N(\mathbf{0}; \mathbf{V}_j \mathbf{K} \mathbf{D}_j \mathbf{K} \boldsymbol{\mu}_j \mathbf{K}, \mathbf{V}_j \mathbf{K})} \quad (11)$$

with $A = \pi(\boldsymbol{\phi}, \boldsymbol{\eta}, \mathbf{K}, \mathbf{t}|\mathbf{d}) \left[\prod_{j=1}^m \prod_{i=1}^{n-1} \pi(v_{ij}|\eta_j, \phi_j, \Delta_i) \right] \left[\prod_{j=1}^m \prod_{i=1}^{n-1} (v_{ij})^{-1/2} \right]$, $\boldsymbol{\mu}_j \mathbf{K} = (\mu_{1jk_1}, \dots, \mu_{njk_n})$, $\mathbf{D}_j \mathbf{K} = \text{diag}(\tau_{1jk_1}, \dots, \tau_{njk_n})$, $\mathbf{V}_j \mathbf{K} = (\mathbf{D}_j \mathbf{K} + \mathbf{W}_j)^{-1}$ and $\mathbf{W}_j = \sum_i v_{ij}^{-1} \mathbf{B}_i \mathbf{B}_i^T$ where \mathbf{B}_i is the i th row of difference matrix \mathbf{B} . Thus the posterior again factorises and we are left with a likelihood containing the parameters of interest $\boldsymbol{\phi}, \boldsymbol{\eta}, \mathbf{v}$. Once inference on these has been performed, we can then obtain $\mathbf{c}|\mathbf{y}, \dots$ from:

$$\pi(\mathbf{c}|\mathbf{y}, \dots) = \int \pi(\mathbf{c}|\mathbf{K}, \mathbf{t}, \mathbf{v}, \mathbf{y}, \mathbf{d}) \pi(\mathbf{K}, \mathbf{t}, \mathbf{v}, \boldsymbol{\eta}|\mathbf{y}, \mathbf{d}) d(\mathbf{K}, \mathbf{t}, \mathbf{v}, \boldsymbol{\phi}, \boldsymbol{\eta})$$

where the first term in the integrand is a normal distribution (given by the left hand side of the numerator of Equation 11) and the second is the posterior from the first inference stage.

We leave the computational details of fitting the model defined in Equation 11 to Technical Appendix B. It suffices to say that the parameters \mathbf{v} require Metropolis-Hastings updates whilst the complete conditionals for $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ are available as Generalised Inverse Gaussian and Gamma distributions respectively. In fact we find it convenient to fix the values of $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ to some suitable prior values. The computation time is substantially reduced because $\mathbf{V} = (\mathbf{D} + \mathbf{W})^{-1}$ is tri-diagonal and thus requires only $\mathcal{O}(n)$ inversion steps. Similarly, for a new proposed value of v_{ij} , say v_{ij}^* , we can write $\mathbf{V}^* = (\mathbf{V}^{-1} + ((v_{ij}^*)^{-1} - v_{ij}^{-1}) \mathbf{B}_i \mathbf{B}_i^T)^{-1}$, where \mathbf{B}_i is the i th row of \mathbf{B} . Inversions of \mathbf{V}^* can be computed via the fast Woodbury formula (e.g. Press et al., 2002). Once this inference stage is completed, we can perform draws of climate \mathbf{c} from the predictive distribution, and subsequently interpolate using the Inverse Gaussian bridge (as in Section 3) to estimate \mathbf{v} at intermediary time points. Interpolations for \mathbf{c} can be found from the resulting Brownian Bridge.

We incorporate time uncertainty in our algorithm by sampling a chronology from $\pi(\mathbf{t}|\mathbf{d})$ at every iteration. This involves another cutting feedback assumption as the parameters of the chronology model, and hence the set of times \mathbf{t} , are informed only by the radiocarbon dates in the core and not by any information on

the sedimentation from the pollen. We feel that this is reasonable though it is an area for further research given that the pollen production of certain varieties of plants or tree surrounding a lake may inform the sedimentation process. The general inclusion of time uncertainty in our model greatly reduces our ability to discern precise volatilities, as can be seen in our case study below. We discuss this property further in Section 7.

In Technical Appendix C, we perform a number of validation steps to discern the performance of our model in ideal and non-ideal simulated situations. As the climate change prior (the evolution equation) is our main focus here and not the likelihood, we use a number of simplified versions of the full likelihood to run tests (though see Appendix C for the full description). We use the following steps to test the algorithm:

1. Fix parameters η_j and ϕ_j for $j = 1, 2, 3$.
2. Generate v_{ij} from the $IG(\eta_j, \phi_j)$ for $i = 1, \dots, n - 1$ and $j = 1, 2, 3$.
3. Generate climates c_{ij} from the evolution equation.
4. Generate ‘psuedo pollen’ from f_θ with θ fixed.

We then perform the steps outlined at the start of this section to produce a joint posterior distribution on $c, v|y$. Note that for simplicity we assume fixed, unit time and we choose $n = 100$ slices which seems reasonable given the types of pollen data we analyse. We evaluate our model by looking at 90% and 50% coverage of the posterior compared to the true values. We perform the above steps 1,000 times in a variety of situations where various model assumptions are violated. In Appendix C we show that, even under quite severe violations, the coverage is still extremely adequate. We perform some further model checking steps as part of the next section.

6 Case study: Sluggan Bog, County Antrim, Northern Ireland

We now apply our model to a site from Northern Ireland, previously published in Smith and Goddard (1991). Our goal is to answer our questions of interest, namely estimating the changing climate and climate volatility over time, to which we have direct access via interpolated values of \mathbf{v} created in stage 5 of our algorithm. We infer the three climate dimensions GDD5 (length of growing season), MTCO (harshness of winter) and AET/PET (availability of moisture), as detailed in Section 2. 28 pollen taxa were used to create MDPs via the methods of Sweeney (2012), with Bchron (Haslett and Parnell, 2008) used to create an age-depth model and thus posterior distributions of the time of each pollen layer. We fix the Inverse Gaussian parameters at the posterior means from the Ice Core data in Section 2, so that $\eta = 2.66$ and $\phi = 15.33$.

The creation of MDPs and their approximation as mixtures is a relatively fast step taking less than 5 minutes on a modern PC with several cores, though the former is strongly dependent on the number of layers n . For our core we have $n = 115$ layers, though other cores may have many more. The MCMC stage was run for 100,000 iterations with a burn-in period of 20,000 and thinning by 40. The resulting 2,000 iterations were checked for convergence using the R package *boa* (Smith, 2005). Posterior creation of climates, and their subsequent interpolation, are of negligible computational impact. A full run of all stages 1-6 took around 15 minutes on an Intel Core-i7 3.4GHz processor with 8 cores and 16Gb of RAM.

Figure 4 shows GDD5, MTCO and AET/PET posterior distributions for Sluggan interpolated on to a regular centurial grid from 0 to 14 ka years BP. A Younger Dryas type event is clearly visibly in MTCO, and there appear to be similar changes in both GDD5 and AET/PET. Unsurprisingly, the last 10k years BP are reasonably constant, much like the ice core data in Figure 2. Figure 5 shows the posterior distributions of interpolated volatilities derived via the Inverse Gaussian bridge. The uncertainties all peak around 12ka BP, corresponding to periods of known rapid climate change. There are similar peaks at about 6ka BP, possibly

corresponding to the start of agricultural activity known as the Neolithic.

It is important to note that the volatilities in Figure 5 are given as 95% credibility intervals. These show that we have large uncertainty about the volatility which is unsurprising given that this is an estimate of a latent process with temporal uncertainty. In Figure 6, we show one of the climate dimensions (MTCO) with time uncertainty ignored. The MDPs are now vertical lines as they no longer exhibit any ‘horizontal’ uncertainty. Unsurprisingly the volatilities are somewhat more clearly defined. In Figure 7, we show MTCO with a Gaussian instead of NIG random walk (equivalent to letting $\phi \rightarrow \infty$). Note that, whilst the large change in climate around 11k years BP is still shown, it is associated with much lower volatility and thus a slower climatic change.

7 Discussion

The model we have presented performs inference on palaeoclimate in a more detailed and thorough fashion than previously possible. The foundation of the model is a state space formulation which explicitly separates out the dynamical system (climate; the evolution equation) from the forward model (the link between climate and proxy pollen data; the state equation). This idea, proposed originally in Haslett et al. (2006), had also been suggested by Tingley et al. (2012). We have implemented and considerably expanded this approach and developed a fast algorithm which can perform inference on both climate and climate volatility through the use of mixtures of marginal data posteriors.

There are several drawbacks to the model as proposed. First, it is conceivable that the mixture indicators do not cover a sufficient variety of tuples \mathbf{K} to sufficiently learn the climate volatility parameters. Such a problem will increase with n and G the number of mixture components used. However, our model validation steps show that this is almost never the case and coverage properties, even when G is under-estimated, still seem adequate. Another disadvantage is the cutting feedback assumptions, first between the likelihood parameters θ and the rest of the model, and second between the chronology model and the NIG process. The former seems most reasonable, as new cores are unlikely to impact much on the climate process given the strength of modern analogue data available. The latter, however, poses an interesting challenge, as if the sedimentation process is also posed as an intrinsic prior it is feasible for inclusion in our MDP-style inferential approach.

The modularity invoked by following our modelling assumptions is of interest in future extensions. This modularity enables various steps to be run in parallel, and also allows us to change modules as required. The steps for the different modules are given at the start of Section 5. For example, to produce the fixed time graph in Figure 6, steps 2 and 3 (creation of MDPs and their approximation as mixtures) did not need to be re-run, being entirely independent of any chronological uncertainty. Thus only the other steps were required. This has larger implications for future modelling as, for example, a new forward model can be used in place of the one we use with no other changes required to any of the other steps. The same applies for the chronology model, the Mclust mixture algorithm, and the evolution equation itself (though it would still need to be intrinsic).

There are several enhancements which follow immediately and to which our algorithm may well still apply. These include: multi-proxy inference, development of new forward models, and spatio-temporal approaches:

- A multi-proxy analysis of palaeoclimate would require multiple forward models describing the relationship between climate and the various proxies. Our state-space MDP approach would not be hindered by such an extension, as we could simply create MDPs for the different proxies and include them as standard, so that the product MDP now becomes, e.g. $\pi_{\text{MDP}}^{\text{proxy1}}(c|y_1) \times \pi_{\text{MDP}}^{\text{proxy2}}(c|y_2)$. However, care needs to be taken in selecting the aspects of climate to which the different proxies are responsive. If these are substantially different, it may be that an extra layer of the state space model is required to match the different climate variables appropriately.

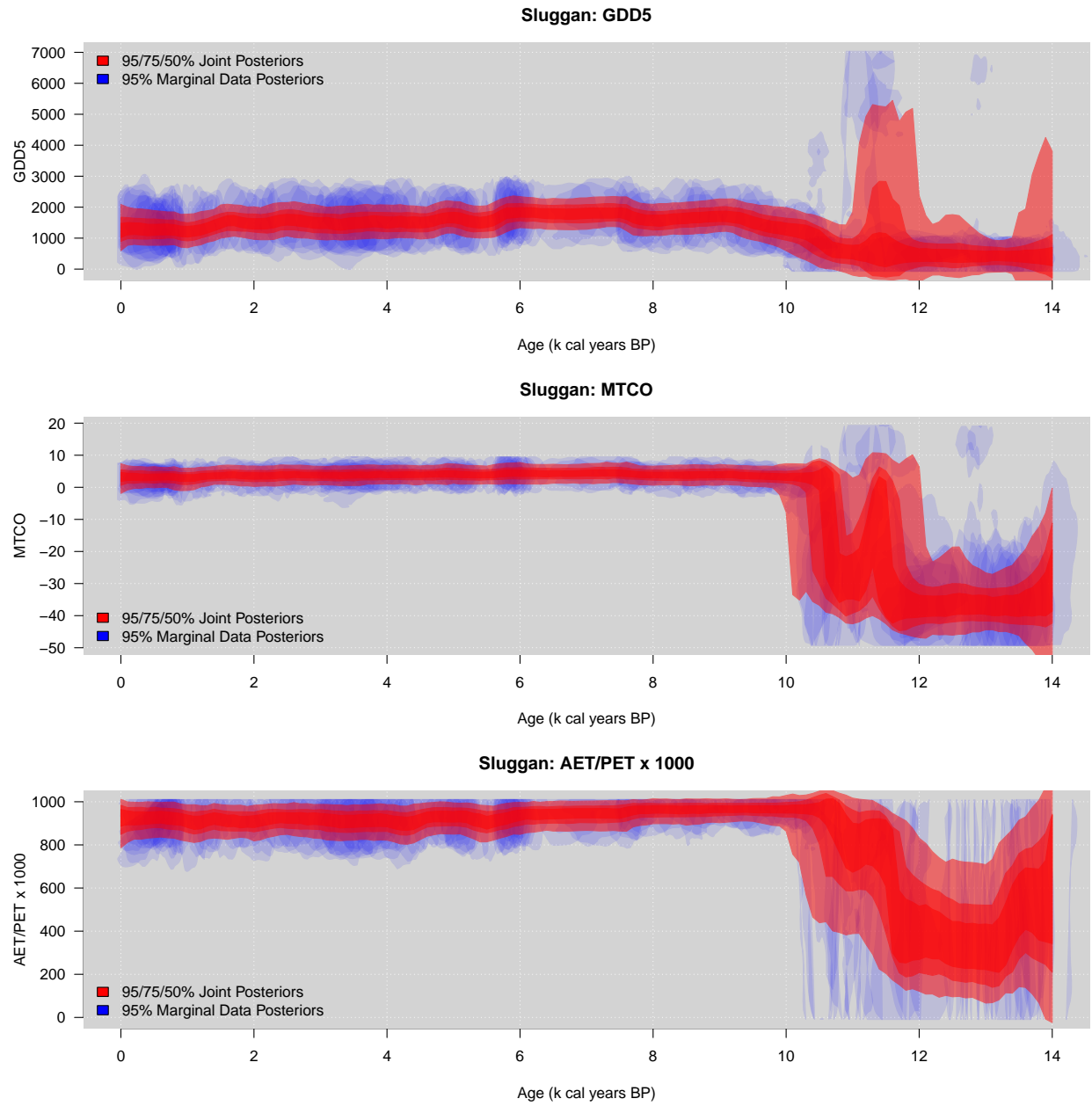


Figure 4: A plot of the centurial interpolated GDD5 (length of growing season), MTCO (harshness of winter) and AET/PET (proportion of available moisture; scaled up to (0,1000)) over the period 0 to 14ka BP. The blue 'blobs' represent the marginal data posteriors whereas the red bands represent summarised posterior interpolations of climates c .

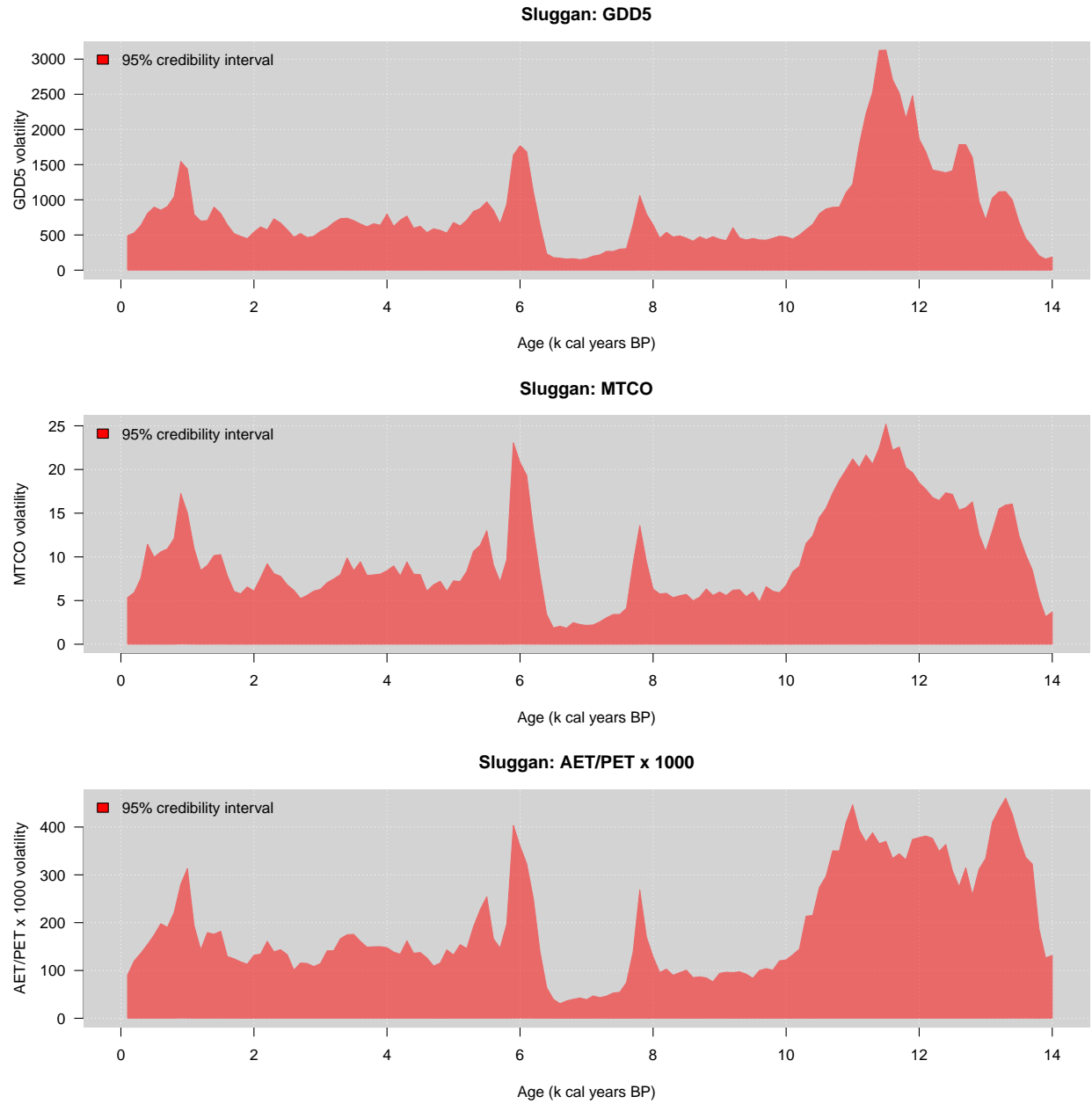


Figure 5: A plot of the posterior centennial interpolated volatilities for the three climate dimensions for Sluggan. The shaded areas represent 95% credibility intervals for the centennial volatility.

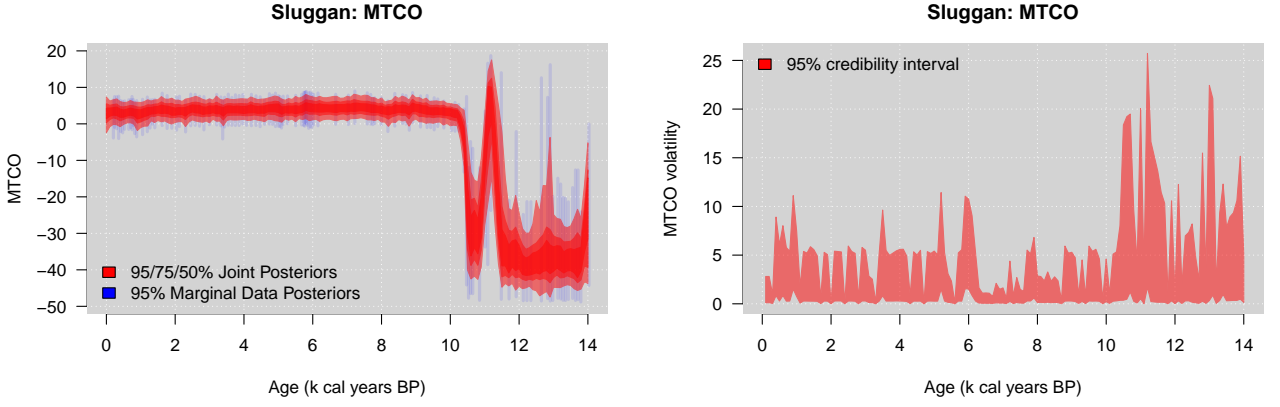


Figure 6: Plots of MTCO (harshness of winter) for the Sluggan core when time is assumed fixed. The left panel shows the posterior distribution of climate c whilst the right panel shows the 95% credible intervals for the centennial interpolated volatility. The legend for the left panel is the same as that of Figure 4.

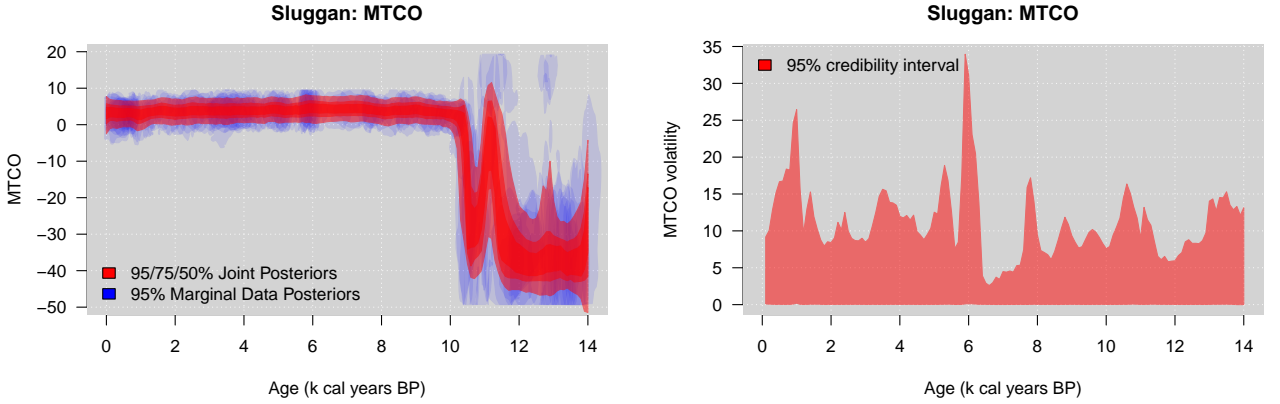


Figure 7: Plots of MTCO (harshness of winter) for the Sluggan core when the evolution equation is assumed to be a Brownian motion. The left panel shows the posterior distribution of climate c whilst the right panel shows the 95% credible intervals for the centennial interpolated volatility. The legend for the left panel is the same as that of Figure 4.

- A key development in palaeoclimate will be the advancement of probabilistic forward models describing the causal chain from climate to proxy data. The forward model we use is relatively simplistic in its description of the mechanics of climate/pollen interaction, though it is far more sophisticated in its description of the uncertainty relating to the counting of pollen data and the relationships between pollen varieties. We encourage the development of physical forward models provided they retain suitable stochastic elements. A recent example of such thinking is Tolwinski-Ward et al. (2011).
- Finally, we might extend our state space approach into the spatial domain and give a richer evolution equation:

$$y(s_i, t_i) | c(s_i, t_i) \sim f_\theta(c(s_i, t_i)), \quad i = 1 \dots, n$$

$$c(s_i, t_i) | \{c(s_1, t_1), \dots, c(s_{i-1}, t_{i-1})\} = c \sim \zeta_\kappa(c), \quad i = 2, \dots, n$$

where now both pollen and climate are indexed by space s and time t and the prior distribution ζ is applied to climate change, parameterised by κ . We might assume that this would use all observations from previous time points t_1, \dots, t_{i-1} so that a particle algorithm might now become more appropriate. The prior ζ might be a simple stochastic climate model, or a richer version of our random walk including covariates and a spatial process. It is immediately obvious that c will no longer factorise out of the posterior, yet if ζ remains intrinsic a Laplace approximation might still allow our algorithm to proceed, though with caveats as to the size of the approximation error. Finally, even in situations where the prior is not intrinsic, it may be that other non-Gaussian mixture arrangements will yield simple tractable forms.

Performing inference on palaeoclimate over multiple sites may be possible by following the recently proposed methodology of Lindgren et al. (2011). In fact, the borrowing of strength from nearby cores may overcome one of our main issues: that of temporal uncertainty. It is certainly feasible that the constrained correlation of neighbouring sites will reduce temporal variability and thus provide more precise estimates of climate and possibly its associated volatility. Following this approach seems most promising in producing a pan-European map of palaeoclimate and its uncertainty.

8 Acknowledgements

We would like to thank the following for fruitful discussions as part of the SUPRAnet network: Jonty Rougier, Caitlin Buck, Tamsin Edwards, and Michel Crucifix. We would also like to thank Brendan Murphy for his assistance with the mclust algorithm.

References

- Barndorff-Nielsen, O. E. (1997, March). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics* 24(1), 1–13.
- Barndorff-Nielsen, O. E. and N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics. *Journal of the Royal Statistical Society - Series B: Statistical Methodology* 63(2), 167–241.
- Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian Computation in population genetics. *Genetics* 162(4), 2025–2035.
- Betrò, B. and R. Rotondi (1991, March). On Bayesian inference for the Inverse Gaussian distribution. *Statistics & Probability Letters* 11(3), 219–224.

- Blaauw, M. and J. A. Christen (2011). Flexible Paleoclimate Age-Depth Models Using an Autoregressive Gamma Process. *Bayesian Analysis* 6(3), 457–474.
- Braconnot, P. and S. Harrison (2003). PMIP (Paleoclimate Modelling Intercomparison Project). *PAGES News* 16(2), 31–32.
- Bronk Ramsey, C. (2008, January). Deposition models for chronological records. *Quaternary Science Reviews* 27(1-2), 42–60.
- Brynjarsdóttir, J. and L. M. Berliner (2011). Bayesian hierarchical modeling for temperature reconstruction from geothermal data. *The Annals of Applied Statistics* 5(2B), 1328–1359.
- Buck, C. E., W. G. Cavanagh, and C. D. Litton (1996). *Bayesian Approach to Interpreting Archaeological Data*. Chichester: John Wiley and Sons Ltd.
- Carvalho, C. M., M. S. Johannes, H. F. Lopes, and N. G. Polson (2010). Particle Learning and Smoothing. *Statistical Science* 25(1), 88–106.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Congdon, P. P. (2001). *Bayesian Statistical Modelling (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section)*. Wiley.
- Davidson, J. (2010, January). Volatility of unevenly sampled fractional Brownian motion: An application to ice core records. *Physical Review E* 81(1).
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological* 39(1), 1–38.
- Doucet, A., E. Arnaudismacjp, and A. M. Johansen (2008). A Tutorial on Particle Filtering and Smoothing : Fifteen years later. *Handbook of Nonlinear Filtering* 1 (December), 4–6.
- Fonseca, T. C. O. and M. F. J. Steel (2011). A general class of nonseparable space-time covariance models. *Environmetrics* 22(May 2009), 224–242.
- Fraley, C. and A. E. Raftery (2002, June). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* 97(458), 611–631.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture And Markov Switching Models*. Springer.
- Garreta, V., P. A. Miller, J. Guiot, C. Hély, S. Brewer, M. T. Sykes, and T. Litt (2009). A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model. *Climate Dynamics* 35(2-3), 371–389.
- Haslett, J. and A. C. Parnell (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society, Series C* 57, 399–418.
- Haslett, J., M. Whitley, S. Bhattacharya, F. J. G. Mitchell, J. R. M. Allen, B. Huntley, S. P. Wilson, and M. Salter-Townshend (2006). Bayesian palaeoclimate reconstruction. *Journal of the Royal Statistical Society, Series A* 169, 395–438.
- Henderson, D. A., R. J. Boys, and D. J. Wilkinson (2010). Bayesian calibration of a stochastic kinetic computer model using multiple data sources. *Biometrics* 66(1), 249–256.
- Holden, P. B., N. R. Edwards, K. I. C. Oliver, T. M. Lenton, and R. D. Wilkinson (2010). A probabilistic calibration of climate sensitivity and terrestrial carbon change in GENIE-1. *Climate Dynamics* 35(5), 785–806.

- Huntley, B. (1993). The use of climate response surfaces to reconstruct palaeoclimate from quaternary pollen and plant macrofossil data. *Philosophical transactions of the royal society of London, series B - Biological sciences*. 341, 215–223.
- Huntley, B. (2012). Reconstructing palaeoclimates from biological proxies: Some often overlooked sources of uncertainty. *Quaternary Science Reviews* 31, 1–16.
- Jansen, E., J. Overpeck, K. R. Briffa, J.-C. Duplessy, F. Joos, V. Masson-Delmotte, D. Olago, B. Otto-Bliesner, W. R. Peltier, S. Rahmstorf, R. Ramesh, D. Raynaud, O. Rind, R. Solomina, V. R., and D. Zhang (2007). Palaeoclimate. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, T. M., and H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Karlis, D. and J. Lillistol (2004, October). Bayesian estimation of NIG models via Markov chain Monte Carlo methods. *Applied Stochastic Models in Business and Industry* 20(4), 323–338.
- Kennedy, M. C. and A. O’Hagan (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society - Series B: Statistical Methodology* 63(3), 425–464.
- Li, B., D. W. Nychka, and C. M. Ammann (2010). The Value of Multiproxy Reconstruction of Past Climate. *Journal of the American Statistical Association* 105(491), 883–895.
- Lindgren, F., H. v. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society - Series B: Statistical Methodology* 73(4), 423–498.
- Mann, M. E., R. S. Bradley, and M. K. Hughes (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* 392(6678), 779–787.
- Mann, M. E., R. S. Bradley, and M. K. Hughes (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters* 26(6), 759.
- Masson-Delmotte, V., M. Kageyama, P. Braconnot, S. Charbit, G. Krinner, C. Ritz, E. Guilyardi, J. Jouzel, A. Abe-Ouchi, M. Crucifix, R. M. Gladstone, C. D. Hewitt, A. Kitoh, A. N. LeGrande, O. Marti, U. Merkel, T. Motoi, R. Ohgaito, B. Otto-Bliesner, W. R. Peltier, I. Ross, P. J. Valdes, G. Vettoretti, S. L. Weber, F. Wolk, and Y. Yu (2006). Past and future polar amplification of climate change: climate model inter-comparisons and ice-core constraints. *Climate Dynamics* 26(5), 513–529.
- McShane, B. B. and A. J. Wyner (2011). Discussion of: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *The Annals of Applied Statistics* 5(1), 1–45.
- Ohlwein, C. and E. R. Wahl (2012). Review of probabilistic pollen-climate transfer methods. *Quaternary Science Reviews* 31, 17–29.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2002). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- R Foundation For Statistical Computing Austria, V. (2011). R Development Core Team, R: a language and environment for statistical computing.
- Ribeiro, C. (2003). A Monte Carlo Method for the Normal Inverse Gaussian Option Valuation Model using an Inverse Gaussian Bridge. *Business* (02), 1–15.
- Salter-Townshend, M. and J. Haslett (2012). Fast Inversion of a Flexible Regression Model for Multivariate, Zero-Inflated Pollen Counts. *Environmetrics In Press*.

- Smith, A. G. and I. C. Goddard (1991). A 12500 Year Record of Vegetational History at Sluggan Bog, Co. Antrim, N. Ireland (Incorporating a Pollen Zone Scheme for the Non-Specialist). *New Phytologist* 118, 167–187.
- Smith, B. J. (2005). Bayesian output analysis program (boa) version 1.1 user’s manual. *Program*, 43.
- Stuiver, M. (2000). GISP2 Oxygen Isotope Ratios. *Quaternary Research* 53(3), 277–284.
- Sweeney, J. (2012). *Advances in Bayesian Model Development and Inversion in Multivariate Inverse Inference Problems with application to palaeoclimate reconstruction*. Ph. D. thesis, Trinity College Dublin.
- Ter Braak, C. (1995). Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. *Chemometrics and Intelligent Laboratory Systems* 28(1), 165–180.
- Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, and B. Rajaratnam (2012, March). Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quaternary Science Reviews* 35, 1–22.
- Tingley, M. P. and P. Huybers (2010). A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part I: Development and Applications to Paleoclimate Reconstruction Problems. *Journal of Climate* 23(10), 2759–2781.
- Tolwinski-Ward, S., M. Evans, M. Hughes, and K. Anchukaitis (2011). An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dynamics* 36(11), 2419–2439.
- Wyse, J., N. Friel, and H. Rue (2010). Approximate simulation-free Bayesian inference for multiple change-point models with dependence within segments. *Bayesian Analysis* 6(4), 24.

A Notation

Below is a table outlining the notation we use throughout the paper and especially in the mathematical derivations contained in Appendix B.

c_{ij}	$= c_j(t_i)$ latent palaeoclimate at time t_i for climate dimension j
\mathbf{c}	vector (length $m \times n$) of latent climate c_{ij} at all times and in all dimensions. Sorted first by time and then by dimension
\mathbf{c}_i	$= \mathbf{c}(t_i)$ vector (length m) of latent palaeoclimate where each element represents each climate dimension at time t_i
\mathbf{c}_j	vector (length n) of latent palaeoclimate where each element represents climate over all time for climate dimension j
d	Depths
f	Forward model
g	group number/mixture component
i	layer number
j	climate dimension number
k	index number
k_i	index number for layer i taking values 1 to G
m	Total number of climate dimensions
n	Total number of layers
o_i	$\delta^{18}\text{O}$ measurement at time t_i
p_{ig}	probability of mixture component g for layer i
t_i	time of layer i in thousands of years before present (ka BP)
v_{ij}	variance of random walk for layer i , climate dimension j
\mathbf{y}	pollen data
\mathbf{y}_i	vector of pollen data for layer i
\mathbf{z}_i	n -vector of zeros with 1 in the i th and -1 in the $(i + 1)$ th position
$\mathbf{D}_j\mathbf{K}$	Diagonal matrix of precisions for climate dimension j under mixture components \mathbf{K}
G	Total number of mixture components
\mathbf{I}_n	Identity matrix of size n
\mathbf{K}	Index set of k_1, k_2, \dots, k_n
$\mathbf{V}_j\mathbf{K}$	Variance matrix for climate dimension j under mixture components \mathbf{K} . $\mathbf{V}_j\mathbf{K} = (\mathbf{D}_j\mathbf{K} + \mathbf{W}_j)^{-1}$
\mathbf{W}_j	Random walk precision matrix for climate dimension j
β	Skew parameter for Inverse Gaussian distribution
Δ_i	Time differences $t_i - t_{i-1}$
η	Parameter of Inverse Gaussian Distribution
θ	Set of parameters governing forward model
μ	Mixture component means, also drift parameter for ice core model
μ_{ijg}	Mixture mean for layer i , climate dimension j , mixture component g
$\boldsymbol{\mu}_{ig}$	m -vector of mixture means for layer i when mixture component g is chosen
$\boldsymbol{\mu}_j\mathbf{K}$	n -vector of mixture means for all layers on climate dimension j under mixture components \mathbf{K}
π	probability distribution
τ	Mixture component precisions
τ_{ijg}	Mixture precision for layer i , climate dimension j , mixture component g
$\boldsymbol{\tau}_{ig}$	$m \times m$ -matrix of mixture precisions for layer i when mixture component g is chosen - often diagonal

B MCMC details

We require posterior simulations of \mathbf{v} via the Metropolis-Hastings algorithm; \mathbf{c} now having been integrated out and hyper-parameters ϕ, η are treated as fixed. The complete conditional for v_{ij} is independent of the other v 's so below we drop the subscripts j for clarity. We update v by proposing new value v^* , and hence create new matrix $\mathbf{Q}_K^* = \mathbf{Q}_K + ((v^*)^{-1} - v^{-1})\mathbf{z}_i\mathbf{z}_i^T$ where $\mathbf{Q}_K = \mathbf{V}_K^{-1}$. Calculate MH ratio as:

$$\begin{aligned}
R_{v_i} &= \frac{N(\mathbf{0}; \mathbf{V}_K \mathbf{D}_K \mu_K, \mathbf{V}_K)}{N(\mathbf{0}; \mathbf{V}_K^* \mathbf{D}_K \mu_K, \mathbf{V}_K^*)} \times \frac{q(v_i|v_i^*)}{q(v_i^*|v_i)} \times \frac{IG(v_i^*|\phi, \eta)}{IG(v_i|\phi, \eta)} \times \frac{(v_i^*)^{-\frac{1}{2}}}{v_i^{-\frac{1}{2}}} \\
&= \frac{|\mathbf{Q}_K|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\mathbf{V}_K \mathbf{D}_K \mu_K)^T \mathbf{Q}_K \mathbf{V}_K \mathbf{D}_K \mu_K]\right\}}{|\mathbf{Q}_K^*|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\mathbf{V}_K^* \mathbf{D}_K \mu_K)^T \mathbf{Q}_K^* \mathbf{V}_K^* \mathbf{D}_K \mu_K]\right\}} \times \frac{q(v_i|v_i^*)}{q(v_i^*|v_i)} \times \frac{IG(v_i^*|\phi, \eta)}{IG(v_i|\phi, \eta)} \times \frac{(v_i^*)^{-\frac{1}{2}}}{v_i^{-\frac{1}{2}}} \\
&= \frac{|\mathbf{Q}_K|^{\frac{1}{2}}}{|\mathbf{Q}_K^*|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[(\mathbf{V}_K \mathbf{D}_K \mu_K)^T \mathbf{Q}_K \mathbf{V}_K \mathbf{D}_K \mu_K - (\mathbf{V}_K^* \mathbf{D}_K \mu_K)^T \mathbf{Q}_K^* \mathbf{V}_K^* \mathbf{D}_K \mu_K]\right\} \times \frac{q(v_i|v_i^*)}{q(v_i^*|v_i)} \times \frac{IG(v_i^*|\phi, \eta)}{IG(v_i|\phi, \eta)} \times \frac{(v_i^*)^{-\frac{1}{2}}}{v_i^{-\frac{1}{2}}} \\
&= \left[\frac{|\mathbf{Q}_K|}{|\mathbf{Q}_K^*|}\right]^{\frac{1}{2}} \exp\left\{-\frac{1}{2}[(\mathbf{D}_K \mu_K)^T (\mathbf{V}_K - \mathbf{V}_K^*) \mathbf{D}_K \mu_K]\right\} \times \frac{q(v_i|v_i^*)}{q(v_i^*|v_i)} \times \frac{IG(v_i^*|\phi, \eta)}{IG(v_i|\phi, \eta)} \times \frac{(v_i^*)^{-\frac{1}{2}}}{v_i^{-\frac{1}{2}}}
\end{aligned}$$

Importantly (dropping the \mathbf{K} subscripts for the time being) $\mathbf{V}^{*-1} = \mathbf{Q}^* = \mathbf{Q} + \mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T$ can be manipulated via the Woodbury formulae:

$$\begin{aligned}
(\mathbf{A} + \mathbf{UCV})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \\
|\mathbf{A} + \mathbf{UCV}| &= |\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U}||\mathbf{C}||\mathbf{A}|
\end{aligned}$$

Thus we have:

$$\begin{aligned}
\frac{|\mathbf{Q}|}{|\mathbf{Q}^*|} &= \frac{|\mathbf{Q}|}{|\mathbf{Q} + \Delta_i^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T|} \\
&= \frac{|\mathbf{Q}|}{|\mathbf{Q}||1 + ((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T \mathbf{V} \mathbf{z}_i|} \\
&= (1 + [(v_i^*)^{-1} - v_i^{-1}]\mathbf{z}_i^T \mathbf{V} \mathbf{z}_i)^{-1}
\end{aligned}$$

Similarly:

$$\begin{aligned}
\mathbf{V} - \mathbf{V}^* = \mathbf{Q}^{-1} - (\mathbf{Q}^*)^{-1} &= \mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i\left([\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T\right)^{-1}\mathbf{z}_i^T \mathbf{Q}^{-1} \\
&= \mathbf{Q}^{-1}\mathbf{z}_i\left([\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T\right)^{-1}\mathbf{z}_i^T \mathbf{Q}^{-1} \\
&= \mathbf{V}\mathbf{z}_i\left([\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T\right)^{-1}\mathbf{z}_i^T \mathbf{V} \\
&= \left([\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T\right)^{-1}\mathbf{V}\mathbf{z}_i\mathbf{z}_i^T \mathbf{V}
\end{aligned}$$

Thus R_{v_i} can be computed from \mathbf{V} without the need for \mathbf{V}^* :

$$R_{v_i} = \left(1 + [(v_i^*)^{-1} - v_i^{-1}]\mathbf{z}_i^T \mathbf{V} \mathbf{z}_i\right)^{-\frac{1}{2}} \exp\left\{-\frac{(\mathbf{D}_K \mu_K)^T \mathbf{V}_K \mathbf{z}_i \mathbf{z}_i^T \mathbf{V}_K \mathbf{D}_K \mu_K}{2\left([\mathbf{Q}^{-1} - \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{z}_i((v_i^*)^{-1} - v_i^{-1})\mathbf{z}_i^T\right)^{-1} + \mathbf{z}_i^T \mathbf{V} \mathbf{z}_i}\right\} \times \frac{q(v_i|v_i^*)}{q(v_i^*|v_i)} \times \frac{IG(v_i^*|\phi, \eta)}{IG(v_i|\phi, \eta)} \times \frac{(v_i^*)^{-\frac{1}{2}}}{v_i^{-\frac{1}{2}}}$$

Note that this involves many repeated calculations of $\mathbf{z}_i^T \mathbf{V} \mathbf{z}_i$ and

$$(\mathbf{D}_K \mu_K)^T \mathbf{V}_K \mathbf{z}_i \mathbf{z}_i^T \mathbf{V}_K \mathbf{D}_K \mu_K = (\mathbf{z}_i^T \mathbf{V}_K \mathbf{D}_K \mu_K)^T \mathbf{z}_i^T \mathbf{V}_K \mathbf{D}_K \mu_K.$$

In R it is possible to use the `limSolve` package and the function `Solve.tridiag` to create \mathbf{V} as it is the inverse of a tri-diagonal matrix. By solving the linear systems we can thus directly obtain, for example $\mathbf{V}_K \mathbf{z}_i$ or $\mathbf{V}_K \mathbf{D}_K \mu_K$, rather than computing \mathbf{V}_K separately.

Scenario	Detail	Proportion inside 90% CI	Proportion inside 50% CI
1	Gaussian likelihood	90.7%	50.8%
2	ZIP likelihood	90.8%	47.7%
3	ZIP likelihood (too few mixture components)	90.1%	44.5%
4a	ZIP likelihood (under-estimated IG parameters)	91.6%	46.5%
4b	ZIP likelihood (over-estimated IG parameters)	94.0%	51.1%

Table 1: Performance of the different model validation scenarios

C Model validation

In this section we determine the properties of our model fitting algorithm using simulated data under some idealised and non-idealised circumstances. To improve the speed of our tests we simplify the likelihood somewhat, though since this is not part of our inference stage in this paper we feel this is reasonable. Similarly we simulate data only observed on fixed, unit time. We consider 5 different scenarios:

1. A simple Gaussian test that the parameters are identifiable when simulated from the model. We set $n = 100$ and $m = 3$. For $j = 1, \dots, m$ we first simulate $\eta_j \sim U(0.1, 10)$ and $\phi_j \sim U(0.1, 10)$. For $i = 1, \dots, n-1$ we then create $v_{ij} \sim IG_2(\mu_j, \phi_j)$ and, for $i = 1, \dots, n$, we create $c_{ij} - c_{i-1,j} \sim N(0, v_{ij})$. Finally we create $\delta_i \sim U(0.02, 2)$ and simulate pseudo pollen $y_{ij} \sim N(c_{ij}, \sqrt{\delta_i^{-1}})$. From the pseudo pollen data and the Gaussian likelihood we obtain Gaussian MDPs (with no simulation or mixture approximation required) which are passed, with the values of ϕ_j and η_j , to our MCMC functions to provide posterior distributions of 3-dimensional climate and volatility.
2. A zero-inflated Poisson likelihood with 3 pollen taxa. The IG parameters, volatilities and climates are simulated as above, but we create pseudo-pollen via $y_{i1} \sim ZIP(p_1, \sqrt{a_1 c_1^2 + a_2 c_2^2})$, $y_{i2} \sim ZIP(p_2, \sqrt{a_1 c_1^2 + a_3 c_3^2})$ and $y_{i3} \sim ZIP(p_3, \sqrt{a_1 c_1^2 + a_2 c_2^2 + a_3 c_3^2})$. Here $ZIP(p, r)$ is a zero-inflated Poisson distribution with zero inflation parameter p and rate r . We set p_1, p_2, p_3 respectively as $p_j \sim U(0, 0.2)$ and a_1, a_2, a_3 as Poisson rate parameters simulated as the modulus of a normal distribution: $a_j \sim |N(0, 1)|$. The pseudo-pollen data are turned into MDPs via importance sampling. The ZIP model gives MDPs that are quite often multi-modal. The MDPs are then approximated as mixtures using $G = 5$ mixture components. These mixture components are then passed to our MCMC algorithm to estimate climates and volatilities.
3. Exactly as (2) but using only 2 mixture components. In many situations this will be a poor representation of the MDP and thus may bias estimates of climate or volatility.
4. Split into two parts:
 - (a) Exactly as (2) but with the Inverse Gaussian parameters η_j and ϕ_j given an underestimating multiplicative bias value simulated from $U(0.5, 1)$.
 - (b) Exactly as (2) but with the Inverse Gaussian parameters η_j and ϕ_j given an overestimating multiplicative bias value simulated from $U(1, 5)$.

We run each of the above 1000 times, and check the coverage properties of the climate posterior to see whether they lie within the 90% and 50% credibility intervals. Table 1 shows the results. Under each scenario the model seems to perform extremely well.

D R Package Bclim

Bclim is available as part of the open source, free, statistical software R (R Foundation For Statistical Computing Austria, 2011). R is available to download from www.r-project.org. To install the package Bclim simply type `install.packages("Bclim")` at the R prompt. The Bclim package is made up of four main functions (covering steps 2 to 6 as given in Section 5), 2 plotting functions for (climate and climate volatility), and a function which runs all necessary steps in sequence.

Example data to run the function can be downloaded from http://mathsci.ucd.ie/~parnell_a/Bclim.html. To run the Sluggan example shown in Section 6, the files should be downloaded via the commands:

```
# Download and load in the response surfaces:
url1 <- 'http://mathsci.ucd.ie/~parnell_a/required.data3D.RData'
download.file(url1,'required_data3D.RData')

# and now the pollen
url2 <- 'http://mathsci.ucd.ie/~parnell_a/SlugganPollen.txt'
download.file(url2,'SlugganPollen.txt')

# and finally the chronologies
url3 <- 'http://mathsci.ucd.ie/~parnell_a/Sluggan_2chrons.txt'
download.file(url3,'Slugganchrons.txt')
```

The response surfaces in the first command above are the pre-calibrated forward model parameters θ . The subsequent functions use the locations of the pollen and chronology file rather than loading them into RAM:

```
# Create variables which state the locations of the pollen and chronologies
pollen.loc <- paste(getwd(),'/SlugganPollen.txt',sep='')
chron.loc <- paste(getwd(),'/Slugganchrons.txt',sep='')

# Load in the response surfaces
load('required.data3D.RData')
```

The functions now proceed as `BclimLayer` which produces the marginal data posteriors, `BclimMixPar` or `BclimMixSer` which approximate the MDPs as mixtures (either in parallel or serial respectively), `BclimMCMC` which produces posterior chains of volatilities and climates, and `BclimInterp` which uses the Inverse Gaussian and Brownian bridges to interpolate climate. Finally `BclimCompile` produces a list object which can be passed to `plotBclim` or `plotBclimVol` for plotting:

```
step1 <- BclimLayer(pollen.loc,required.data3D=required.data3D)
step2 <- BclimMixPar(step1)
step3 <- BclimMCMC(step2,chron.loc)
step4 <- BclimInterp(step2,step3)
results <- BclimCompile(step1,step2,step3,step4,core.name="Sluggan")

# Create a plot of MTCO (dim=2)
plotBclim(results,dim=2)

# Create a volatility plot
plotBclimVol(results,dim=2)
```

Each of the above functions has an associated help file which provides further information and options.